

RECEIVED  
FEB 25 2002  
TECH CENTER 1600/2900



1651

## RAW SEQUENCE LISTING

DATE: 02/13/2002

PATENT APPLICATION: US/09/930,440B

TIME: 15:24:22

Input Set : A:\PF509p2SecondSubSeqList.txt

Output Set: N:\CRF3\02132002\I930440B.raw

PS

ENTERED

3 <110> APPLICANT: Betenbaugh et al.  
 5 <120> TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways  
 7 <130> FILE REFERENCE: PF509P2  
 9 <140> CURRENT APPLICATION NUMBER: 09/930,440B  
 10 <141> CURRENT FILING DATE: 2001-08-16  
 12 <150> PRIOR APPLICATION NUMBER: 60/227,579  
 13 <151> PRIOR FILING DATE: 2000-08-25  
 15 <150> PRIOR APPLICATION NUMBER: 09/516,793  
 16 <151> PRIOR FILING DATE: 2000-03-01  
 18 <150> PRIOR APPLICATION NUMBER: 60/169,624  
 19 <151> PRIOR FILING DATE: 1999-12-08  
 21 <150> PRIOR APPLICATION NUMBER: 60/122,582  
 22 <151> PRIOR FILING DATE: 1999-03-02  
 24 <160> NUMBER OF SEQ ID NOS: 8  
 26 <170> SOFTWARE: PatentIn Ver. 2.1  
 28 <210> SEQ ID NO: 1  
 29 <211> LENGTH: 1429  
 30 <212> TYPE: DNA  
 31 <213> ORGANISM: Homo sapiens  
 33 <220> FEATURE:  
 34 <221> NAME/KEY: CDS  
 35 <222> LOCATION: (1)..(693)  
 37 <400> SEQUENCE: 1  
 38 atg gcc ttc cca aag aag aaa ctt cag ggt ctt gtg gct gca acc atc 48  
 39 Met Ala Phe Pro Lys Lys Lys Leu Gln Gly Leu Val Ala Ala Thr Ile  
 40 1 5 10 15  
 42 acg cca atg act gag aat gga gaa atc aac ttt tca gta att ggt cag 96  
 43 Thr Pro Met Thr Glu Asn Gly Glu Ile Asn Phe Ser Val Ile Gly Gln  
 44 20 25 30  
 46 tat gtg gat tat ctt gtg aaa gaa cag gga gtg aag aac att ttt gtg 144  
 47 Tyr Val Asp Tyr Leu Val Lys Glu Gln Gly Val Lys Asn Ile Phe Val  
 48 35 40 45  
 50 aat ggc aca aca gga gaa ggc ctg tcc ctg agc gtc tca gag cgt cgc 192  
 51 Asn Gly Thr Thr Gly Glu Gly Leu Ser Leu Ser Val Ser Glu Arg Arg  
 52 50 55 60  
 54 cag gtt gca gag gag tgg gtg aca aaa ggg aag gac aag ctg gat cag 240  
 55 Gln Val Ala Glu Glu Trp Val Thr Lys Gly Lys Asp Lys Leu Asp Gln  
 56 65 70 75 80  
 58 gtg ata att cac gta gga gca ctg agc ttg aag gag tca cag gaa ctg 288  
 59 Val Ile Ile His Val Gly Ala Leu Ser Leu Lys Glu Ser Gln Glu Leu  
 60 85 90 95  
 62 gcc caa cat gca gca gaa ata gga gct gat ggc atc gct gtc att gca 336  
 63 Ala Gln His Ala Ala Glu Ile Gly Ala Asp Gly Ile Ala Val Ile Ala

RECEIVED  
FEB 25 2002  
TECH CENTER 1600/2900

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/930,440B

DATE: 02/13/2002  
TIME: 15:24:22

Input Set : A:\PF509p2SecondSubSeqList.txt  
Output Set: N:\CRF3\02132002\I930440B.raw

```

64          100          105          110
66 ccg ttc ttc ctc aag cca tgg acc aaa gat atc ctg att aat ttc cta 384
67 Pro Phe Phe Leu Lys Pro Trp Thr Lys Asp Ile Leu Ile Asn Phe Leu
68          115          120          125
70 aag gaa gtg gct gct gcc gcc cct gcc ctg cca ttt tat tac tat cac 432
71 Lys Glu Val Ala Ala Ala Pro Ala Leu Pro Phe Tyr Tyr Tyr His
72          130          135          140
74 att cct gcc ttg aca ggg gta aag att cgt gct gag gag ttg ttg gat 480
75 Ile Pro Ala Leu Thr Gly Val Lys Ile Arg Ala Glu Glu Leu Leu Asp
76 145          150          155          160
78 ggg att ctg gat aag atc ccc acc ttc caa ggg ctg aaa ttc agt gat 528
79 Gly Ile Leu Asp Lys Ile Pro Thr Phe Gln Gly Leu Lys Phe Ser Asp
80          165          170          175
82 aca gat ctc tta gac ttc ggg caa tgt gtt gat cag aat cgc cag caa 576
83 Thr Asp Leu Leu Asp Phe Gly Gln Cys Val Asp Gln Asn Arg Gln Gln
84          180          185          190
86 cag ttt gct ttc ctt ttt ggg gtg gat gag caa ctg ttg agt gct ctg 624
87 Gln Phe Ala Phe Leu Phe Gly Val Asp Glu Gln Leu Leu Ser Ala Leu
88          195          200          205
90 gtg atg gga gca act gga gca gtg ggc agt ttt gta tcc aga gat tta 672
91 Val Met Gly Ala Thr Gly Ala Val Gly Ser Phe Val Ser Arg Asp Leu
92          210          215          220
94 tca act ttg ttg tca aac tag gttttggagt gtcacagacc aaagccatca 723
95 Ser Thr Leu Leu Ser Asn
96 225          230
98 tgactctggt ctctgggatt ccaatgggcc caccocggct tccactgcag aaagcctcca 783
100 gggagtttac tgatagtgt gaagctaaac tgaagagcct ggatttcctt tctttcactg 843
102 atttaaagga tggaaacttg gaagctggta gctagtgcct ctctatcaaa tcaggggttg 903
104 caccttgaga cataatctac cttaaatagt gcattttttt ctcaggggaat tttagatgaa 963
106 cttgaataaa ctctcctagc aaatgaaatc tcacaataag cattgaggta ccttttgta 1023
108 gccttaaaaa gtcttatttt gtgaaggggc aaaaactcta ggagtcacaa ctctcagtca 1083
110 ttcatttcac agattttttt gtggagaaat ttctgtttat atggatgaaa tggaaatcaag 1143
112 aggaaaattg taattgatta attccatctg tctttaggag ctctcattat ctcggtctct 1203
114 ggttcctaatt cctatttttaa agttgtctaa ttttaaacca ctataatatg tcttcatttt 1263
116 aataaatatt catttggaat ctaggaaaac tctgagctac tgcatttagg caggcacttt 1323
118 aataccaaac tgtaacatgt ctcaactgta tacaactcaa aatacaccag ctcatgtggc 1383
120 tgctcagctt aactctagaa tggatgcttt tgaattcatt tcgatg 1429
123 <210> SEQ ID NO: 2
124 <211> LENGTH: 230
125 <212> TYPE: PRT
126 <213> ORGANISM: Homo sapiens
128 <400> SEQUENCE: 2
129 Met Ala Phe Pro Lys Lys Lys Leu Gln Gly Leu Val Ala Ala Thr Ile
130 1 5 10 15
131 Thr Pro Met Thr Glu Asn Gly Glu Ile Asn Phe Ser Val Ile Gly Gln
132 20 25 30
133 Tyr Val Asp Tyr Leu Val Lys Glu Gly Val Lys Asn Ile Phe Val
134 35 40 45
135 Asn Gly Thr Thr Gly Glu Gly Leu Ser Leu Ser Val Ser Glu Arg Arg

```

## RAW SEQUENCE LISTING

DATE: 02/13/2002

PATENT APPLICATION: US/09/930,440B

TIME: 15:24:22

Input Set : A:\PF509p2SecondSubSeqList.txt

Output Set: N:\CRF3\02132002\I930440B.raw

```

136      50      55      60
137 Gln Val Ala Glu Glu Trp Val Thr Lys Gly Lys Asp Lys Leu Asp Gln
138 65      70      75      80
139 Val Ile Ile His Val Gly Ala Leu Ser Leu Lys Glu Ser Gln Glu Leu
140      85      90      95
141 Ala Gln His Ala Ala Glu Ile Gly Ala Asp Gly Ile Ala Val Ile Ala
142      100      105      110
143 Pro Phe Phe Leu Lys Pro Trp Thr Lys Asp Ile Leu Ile Asn Phe Leu
144      115      120      125
145 Lys Glu Val Ala Ala Ala Ala Pro Ala Leu Pro Phe Tyr Tyr Tyr His
146      130      135      140
147 Ile Pro Ala Leu Thr Gly Val Lys Ile Arg Ala Glu Glu Leu Leu Asp
148 145      150      155      160
149 Gly Ile Leu Asp Lys Ile Pro Thr Phe Gln Gly Leu Lys Phe Ser Asp
150      165      170      175
151 Thr Asp Leu Leu Asp Phe Gly Gln Cys Val Asp Gln Asn Arg Gln Gln
152      180      185      190
153 Gln Phe Ala Phe Leu Phe Gly Val Asp Glu Gln Leu Leu Ser Ala Leu
154      195      200      205
155 Val Met Gly Ala Thr Gly Ala Val Gly Ser Phe Val Ser Arg Asp Leu
156      210      215      220
157 Ser Thr Leu Leu Ser Asn
158 225      230

```

162 &lt;210&gt; SEQ ID NO: 3

163 &lt;211&gt; LENGTH: 1305

164 &lt;212&gt; TYPE: DNA

165 &lt;213&gt; ORGANISM: Homo sapiens

167 &lt;220&gt; FEATURE:

168 &lt;221&gt; NAME/KEY: CDS

169 &lt;222&gt; LOCATION: (1)..(1305)

171 &lt;220&gt; FEATURE:

172 &lt;221&gt; NAME/KEY: misc\_feature

173 &lt;222&gt; LOCATION: (397)..(399)

174 &lt;223&gt; OTHER INFORMATION: The 'yat' at location 397..399 encodes amino acid His, or

Tyr.

176 &lt;220&gt; FEATURE:

177 &lt;221&gt; NAME/KEY: misc\_feature

178 &lt;222&gt; LOCATION: (406)..(408)

179 &lt;223&gt; OTHER INFORMATION: The 'gkt' at location 406..408 encodes amino acid Gly, or

Val.

181 &lt;220&gt; FEATURE:

182 &lt;221&gt; NAME/KEY: misc\_feature

183 &lt;222&gt; LOCATION: (439)..(441)

184 &lt;223&gt; OTHER INFORMATION: The 'yca' at location 439..441 encodes amino acid Pro, or

Ser.

186 &lt;220&gt; FEATURE:

187 &lt;221&gt; NAME/KEY: misc\_feature

188 &lt;222&gt; LOCATION: (505)..(507)

189 &lt;223&gt; OTHER INFORMATION: The 'gkt' at location 505..507 encodes amino acid Gly, or

Val.

191 &lt;400&gt; SEQUENCE: 3

192 atg gac tcg gtg gag aag ggg gcc gcc acc tcc gtc tcc aac ccg cgg 48

193 Met Asp Ser Val Glu Lys Gly Ala Ala Thr Ser Val Ser Asn Pro Arg

## RAW SEQUENCE LISTING

DATE: 02/13/2002

PATENT APPLICATION: US/09/930,440B

TIME: 15:24:22

Input Set : A:\PF509p2SecondSubSeqList.txt

Output Set: N:\CRF3\02132002\I930440B.raw

194 1 5 10 15  
 196 ggg cga ccg tcc cgg ggc cgg ccg ccg aag ctg cag cgc aac tct cgc 96  
 197 Gly Arg Pro Ser Arg Gly Arg Pro Pro Lys Leu Gln Arg Asn Ser Arg  
 198 20 25 30  
 200 ggc ggc cag ggc cga ggt gtg gag aag ccc ccg cac ctg gca gcc cta 144  
 201 Gly Gly Gln Gly Arg Gly Val Glu Lys Pro Pro His Leu Ala Ala Leu  
 202 35 40 45  
 204 att ctg gcc cgg gga ggc agc aaa ggc atc ccc ctg aag aac att aag 192  
 205 Ile Leu Ala Arg Gly Gly Ser Lys Gly Ile Pro Leu Lys Asn Ile Lys  
 206 50 55 60  
 208 cac ctg gcg ggg gtc ccg ctc att ggc tgg gtc ctg cgt gcg gcc ctg 240  
 209 His Leu Ala Gly Val Pro Leu Ile Gly Trp Val Leu Arg Ala Ala Leu  
 210 65 70 75 80  
 212 gat tca ggg gcc ttc cag agt gta tgg gtt tcc aca gac cat gat gaa 288  
 213 Asp Ser Gly Ala Phe Gln Ser Val Trp Val Ser Thr Asp His Asp Glu  
 214 85 90 95  
 216 att gag aat gtg gcc aaa caa ttt ggt gca caa gtt cat cga aga agt 336  
 217 Ile Glu Asn Val Ala Lys Gln Phe Gly Ala Gln Val His Arg Arg Ser  
 218 100 105 110  
 220 tct gaa gtt tca aaa gac agc tct acc tca cta gat gcc atc ata gaa 384  
 221 Ser Glu Val Ser Lys Asp Ser Ser Thr Ser Leu Asp Ala Ile Ile Glu  
 222 115 120 125  
 224 ttt ctt aat tat yat aat gag gkt gac att gta gga aat att caa gct 432  
 W--> 225 Phe Leu Asn Tyr Xaa Asn Glu Xaa Asp Ile Val Gly Asn Ile Gln Ala  
 226 130 135 140  
 228 act tct yca tgt tta cat cct act gat ctt caa aaa gtt gca gaa atg 480  
 W--> 229 Thr Ser Xaa Cys Leu His Pro Thr Asp Leu Gln Lys Val Ala Glu Met  
 230 145 150 155 160  
 232 att cga gaa gaa gga tat gat tct gkt ttc tct gtt gtg aga cgc cat 528  
 W--> 233 Ile Arg Glu Glu Gly Tyr Asp Ser Xaa Phe Ser Val Val Arg Arg His  
 234 165 170 175  
 236 cag ttt cga tgg agt gaa att cag aaa gga gtt cgt gaa gtg acc gaa 576  
 237 Gln Phe Arg Trp Ser Glu Ile Gln Lys Gly Val Arg Glu Val Thr Glu  
 238 180 185 190  
 240 cct ctg aat tta aat cca gct aaa cgg cct cgt cga caa gac tgg gat 624  
 241 Pro Leu Asn Leu Asn Pro Ala Lys Arg Pro Arg Arg Gln Asp Trp Asp  
 242 195 200 205  
 244 gga gaa tta tat gaa aat ggc tca ttt tat ttt gct aaa aga cat ttg 672  
 245 Gly Glu Leu Tyr Glu Asn Gly Ser Phe Tyr Phe Ala Lys Arg His Leu  
 246 210 215 220  
 248 ata gag atg ggt tac ttg cag ggt gga aaa tgg cat act acg aaa tgc 720  
 249 Ile Glu Met Gly Tyr Leu Gln Gly Gly Lys Trp His Thr Thr Lys Cys  
 250 225 230 235 240  
 252 gag ctg gaa cat agt gtg gat ata gat gtg gat att gat tgg cct att 768  
 253 Glu Leu Glu His Ser Val Asp Ile Asp Val Asp Ile Asp Trp Pro Ile  
 254 245 250 255  
 256 gca gag caa aga gta tta aga tat ggc tat ttt ggc aaa gag aag ctt 816  
 257 Ala Glu Gln Arg Val Leu Arg Tyr Gly Tyr Phe Gly Lys Glu Lys Leu  
 258 260 265 270

## RAW SEQUENCE LISTING

DATE: 02/13/2002

PATENT APPLICATION: US/09/930,440B

TIME: 15:24:22

Input Set : A:\PF509p2SecondSubSeqList.txt

Output Set: N:\CRF3\02132002\I930440B.raw

```

260 aag gaa ata aaa ctt ttg gtt tgc aat att gat gga tgt ctc acc aat      864
261 Lys Glu Ile Lys Leu Leu Val Cys Asn Ile Asp Gly Cys Leu Thr Asn
262      275      280      285
264 ggc cac att tat gta tca gga gac caa aaa gaa ata ata tct tat gat      912
265 Gly His Ile Tyr Val Ser Gly Asp Gln Lys Glu Ile Ile Ser Tyr Asp
266      290      295      300
268 gta aaa gat gct att ggg ata agt tta tta aag aaa agt ggt att gag      960
269 Val Lys Asp Ala Ile Gly Ile Ser Leu Leu Lys Lys Ser Gly Ile Glu
270 305      310      315      320
272 gtg agg cta atc tca gaa agg gcc tgt tca aag cag acg ctg tct tct      1008
273 Val Arg Leu Ile Ser Glu Arg Ala Cys Ser Lys Gln Thr Leu Ser Ser
274      325      330      335
276 tta aaa ctg gat tgc aaa atg gaa gtc agt gta tca gac aag cta gca      1056
277 Leu Lys Leu Asp Cys Lys Met Glu Val Ser Val Ser Asp Lys Leu Ala
278      340      345      350
280 gtt gta gat gaa tgg aga aaa gaa atg ggc ctg tgc tgg aaa gaa gtg      1104
281 Val Val Asp Glu Trp Arg Lys Glu Met Gly Leu Cys Trp Lys Glu Val
282      355      360      365
284 gca tat ctt gga aat gaa gtg tct gat gaa gag tgc ttg aag aga gtg      1152
285 Ala Tyr Leu Gly Asn Glu Val Ser Asp Glu Glu Cys Leu Lys Arg Val
286      370      375      380
288 ggc cta agt ggc gct cct gct gat gcc tgt tcc tac gcc cag aag gct      1200
289 Gly Leu Ser Gly Ala Pro Ala Asp Ala Cys Ser Tyr Ala Gln Lys Ala
290 385      390      395      400
292 gtt gga tac att tgc aaa tgt aat ggt ggc cgt ggt gcc atc cga gaa      1248
293 Val Gly Tyr Ile Cys Lys Cys Asn Gly Gly Arg Gly Ala Ile Arg Glu
294      405      410      415
296 ttt gca gag cac att tgc cta cta atg gaa aaa gtt aat aat tca tgc      1296
297 Phe Ala Glu His Ile Cys Leu Leu Met Glu Lys Val Asn Asn Ser Cys
298      420      425      430
300 caa aaa tag      1305
301 Gln Lys
304 <210> SEQ ID NO: 4
305 <211> LENGTH: 434
306 <212> TYPE: PRT
307 <213> ORGANISM: Homo sapiens
309 <220> FEATURE:
310 <221> NAME/KEY: misc_feature
311 <222> LOCATION: (133)..(133)
312 <223> OTHER INFORMATION: The 'Xaa' at location 133 stands for His, or Tyr.
314 <220> FEATURE:
315 <221> NAME/KEY: misc_feature
316 <222> LOCATION: (136)..(136)
317 <223> OTHER INFORMATION: The 'Xaa' at location 136 stands for Gly, or Val.
319 <220> FEATURE:
320 <221> NAME/KEY: misc_feature
321 <222> LOCATION: (147)..(147)
322 <223> OTHER INFORMATION: The 'Xaa' at location 147 stands for Pro, or Ser.
324 <220> FEATURE:

```

Use of n and/or Xaa has been detected in the Sequence Listing.  
 Review the Sequence Listing to insure a corresponding  
 explanation is presented in the <220> to <223> fields of  
 each sequence using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/930,440B

DATE: 02/13/2002

TIME: 15:24:23

Input Set : A:\PF509p2SecondSubSeqList.txt

Output Set: N:\CRF3\02132002\I930440B.raw

L:225 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:229 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:233 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:346 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4  
L:348 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4  
L:350 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4